数理解析研究所講究録
571巻 1985年 111-121

Transitive points under the modular group and continued fractions

Shunsuke Morosawa (諸沢俊介)
Department of Mathematics, Tôhoku University

0. Introduction.

A Fuchsian group is a discrete subgroup of linear fractional transformations each of which preserves a unit disk $D = \{z \mid |z|<1\}$ (or upper half plane $H = \{z = x+iy \mid y>0\}$). Denote the boundary of D and of H by S and $\hat{R}$ respectively. Since a Fuchsian group acting on H is conjugate to some Fuchsian group acting on D by some linear fractional transformation, we consider a Fuchsian group acting on D or H case by case. We think that D and H are both equiped with Poincaré metric. The ergodic properties of Fuchsian groups have been investigated by many authors (e.g. [2],[6]). In this paper, we consider the following property. Let Γ be a Fuchsian group acting on D. We call a point $\zeta \in S$ is a transitive point under Γ if, for all ordered pair (ζ_1, ζ_2) of two distinct points of S and all $z \in D$ and for all $\varepsilon>0$, there exists an element $\gamma \in \Gamma$ such that $|\zeta_1-\gamma(z)|+|\zeta_2-\gamma(\zeta)|<\varepsilon$. In fact, the transitivity is independent of the choice of z (see [4]). The transitivity associated to a Fuchsian group acting on H is defined similarly. If ζ is not a transitive point, we call it an intransitive point under Γ. For example, parabolic fixed points of Γ are intransitive points under Γ (see [4]). In [5], it showed that if Γ is a finitely generated Fuchsian group of the first kind, almost all points of S are transitive points under Γ. But what points are transitive under Γ? We consider this problem in the case of the modular group G. In this case, Artin [1] investigated the transitivity of geodesic lines as **Quasiergodizität**. The modular group is a Fuchsian group acting on H and each of whose elements is of the form

$$g(z) = \frac{az+b}{cz+d} \quad a,b,c, \text{ and } d \text{ are integers } \& \quad ad-bc = 1.$$

By $[n_0, n_1, n_2,\cdots]$ we denote the continued fraction $x = n_0 + \frac{1}{n_1} + \frac{1}{n_2} + \cdots$, where n_0 is non-negative integer and n_i, $i \geq 0$, is a positive integer. If $x<0$, we define $x = -[n_0, n_1, n_2 \cdots]$ for $-x = n_0 + \frac{1}{n_1} + \frac{1}{n_2} + \cdots$. All rational numbers are parabolic fixed points of G, so we consider only irrational numbers. Using continued fractions, we give a characterization of transitive points under the modular group, as follows.

Theorem 1. A point $x = \pm[n_0, n_1, n_2,\cdots]$ is a transitive point under the modular group G if and only if, for an arbitrary finite sequence $a_0, a_1,\cdots a_m$, where a_i is a positive integer, there exists k such that $n_k = a_0, n_{k+1} = a_1, \cdots n_{k+m} = a_m$.

In §1, we prove some lemmas and a theorem on transitive points under an arbitrary Fuchsian group. In §2, we shortly explain the cutting sequence which is defined in [7] by Series. In §3, we give the proof of Theorem 1. In §4, using the cutting sequence, we give another proof of a certain theorem concerning to continued fractions.

1. Theorem on transitivity.

Let ζ, ζ_1 and ζ_2 be in S and let z be in D. By $L(\zeta_1,\zeta_2)$ and $R(z,\zeta)$ we denote the oriented geodesic line whose initial point ζ_1 and whose terminal point ζ_2 and the oriented geodesic ray whose initial point z and whose terminal point ζ respectively. We say that $R(z,\zeta)$ (or $L(\zeta_1,\zeta_2)$) converges to $L(\theta_1,\theta_2)$ if, for all $\varepsilon>0$, there exists an element $\gamma\epsilon\Gamma$ such that $|\gamma(z)-\theta_1|+|\gamma(\zeta)-\theta_2|<\varepsilon$ (or $|\gamma(\zeta_1)-\theta_1|+|\gamma(\zeta_2)-\theta_2|$ $<\varepsilon$). Using this notation, we say the definition of transitivity in §0 as follows. A point ζ is called a transitive point under Γ if, for an arbitrary geodesic line $L(\theta_1,\theta_2)$ and an arbitrary geodesic ray $R(z,\zeta)$, $R(z,\zeta)$ converges to $L(\theta_1,\theta_2)$. We also say that $L(\zeta_1,\zeta_2)$ is a transitive geodesic line under Γ if, for an arbitrary geodesic line $L(\theta_1,\theta_2)$, the geodesic line $L(\zeta_1,\zeta_2)$ converges to $L(\theta_1,\theta_2)$. If $L(\zeta_1,\zeta_2)$ is not a transitive geodesic line under Γ, we call it an intransitive geodesic line under Γ. For example, let ζ_1 and ζ_2 be fixed points of a hyperbolic element of Γ. Then $L(\zeta_1,\zeta_2)$ is an intransitive geodesic line.

In this section, we assume that Γ is an arbitrary Fuchsian group, but not an elementary group. Hence Γ has hyperbolic elements. Let ζ_1 and ζ_2 be fixed points of a hyperbolic element of Γ. The geodesic ray $R(z,\zeta_2)$ converges to only

$L(\zeta_1,\zeta_2)$ and its Γ-images. Thus ζ_2 is an intransitive point. Similarly all the hyperbolic fixed points of Γ are intransitive points under Γ.

In the proofs of the following lemmas and a theorem, we only consider Fuchsian groups acting on D. But the result is true for Fuchsian groups acting on H.

Lemma 1. Let $L(\zeta_1,\zeta_2)$ be an intransitive geodesic line under Γ. If $L(\zeta_1,\zeta_2)$ converges to some $L(\theta_1,\theta_2)$, then $L(\theta_1,\theta_2)$ is an intransitive geodesic line under Γ.

Proof. Assume that $L(\theta_1,\theta_2)$ is a transitive geodesic line. Then, for an arbitrary geodesic line $L(\eta_1,\eta_2)$ and for all $\varepsilon>0$, there exists an element $\gamma\in\Gamma$ such that $|\gamma(\theta_1)-\eta_1|+|\gamma(\theta_2)-\eta_2|<\varepsilon$. Since each element of Γ maps S to S continuously, there exists $\delta>0$ such that, for $|\theta_1-\theta_1'|+|\theta_2-\theta_2'|<\delta$, $|\gamma(\theta_1')-\eta_1| +|\gamma(\theta_2')-\eta_2|<\varepsilon$. For this δ, there exists an element $\beta\in\Gamma$ such that $|\beta(\zeta_1)-\theta_1|+|\beta(\zeta_2)-\theta_2|<\delta$ since $L(\zeta_1,\zeta_2)$ converges to $L(\theta_1,\theta_2)$. Hence we have $|\gamma\beta(\zeta_1)-\eta_1|+|\gamma\beta(\zeta_2)-\eta_2|<\varepsilon$. This shows that $L(\zeta_1,\zeta_2)$ converges to an arbitrary geodesic line $L(\eta_1,\eta_2)$. This contradicts the assumption of $L(\zeta_1,\zeta_2)$. q.e.d.

Lemma 2. If the geodesic ray $R(z,\zeta)$ converges to an arbitrary transitive geodesic line, then ζ is a transitive point under Γ.

Proof. We take a transitive geodesic line $L(\theta_1,\theta_2)$. For an arbitrary geodesic line $L(\eta_1,\eta_2)$ and all $\varepsilon>0$, there exists an element $\gamma\in\Gamma$ such that $|\gamma(\theta_1)-\eta_1|+|\gamma(\theta_2)-\eta_2|<\varepsilon/2$. Since $L(\gamma(\theta_1),\gamma(\theta_2))$ is also a transitive geodesic line, there exists an element $\beta\in\Gamma$ such that $|\beta(z)-\gamma(\theta_1)|+|\beta(\zeta)-\gamma(\theta_2)|<\varepsilon/2$. Hence we have $|\beta(z)-\eta_1|+|\beta(\zeta)-\eta_2|<\varepsilon$. This shows that ζ is a transitive point under Γ. q.e.d.

Using above two lemmas, we prove the following theorem.

Theorem 2. Both ζ_1 and ζ_2 are intransitive points under Γ if and only if $L(\zeta_1,\zeta_2)$ is an intransitive geodesic line under Γ.

Proof. The sufficient condition is clear from the definitions.

First, we assume that at least one of ζ_1 and ζ_2 is a hyperbolic fixed point of Γ, say ζ_1. By ζ_1' we denote another fixed point of the hyperbolic element which fixes ζ_1. We take $z\in L(\zeta_1,\zeta_2)\cap D$. The geodesic ray $R(z,\zeta_1)$ converges to only $L(\zeta_1',\zeta_1)$ and its Γ-images. Hence $R(z,\zeta_2)$ must converges to an arbitrary transitive geodesic line, if $L(\zeta_1,\zeta_2)$ is a transitive geodesic line. Therefore, by Lemma 2, ζ_2 is a transitive point. This is contradiction. Hence $L(\zeta_1,\zeta_2)$ is an intransitive geodesic line.

Next, we assume that neither ζ_1 nor ζ_2 is a hyperbolic fixed point. Take an arbitrary hyperbolic fixed point ζ_3. By the above argument, $L(\zeta_3,\zeta_2)$ is an intransitive geodesic line. So, by Lemma 1, $R(z,\zeta_2)$ converges to only intransitive geodesic lines. Hence $R(z,\zeta_1)$ converges to an arbitrary transitive geodesic line, if $L(\zeta_1,\zeta_2)$ is a transitive geodesic line. This means ζ_1 is a transitive point by Lemma 2. Therefor $L(\zeta_1,\zeta_2)$ is an intransitive geodesic line. q.e.d.

2. Cutting sequences.

In the following sections, by G we denote the modular group. We consider the Farey tesselation F, the tesselation of H by images of the imaginary axis under G. Each tessera of F is a non-euclidean triangle whose vertices are all on $\hat{R}$. An arbitrary oriented geodesic line $L(x,y)$ is divided into oriented segments by the triangles of F. We label each oriented segment either R or L according as two sides of the triangle which the segment crosses meet to the right or left of the segment. If $L(x,y)$ starts from a vertex of some triangle or ends in a vertex of some triangle, we may label the segment R or L freely. We arrange the letters R and L as according to the order of the directed segments of $L(x,y)$. If R's (or L's) are succeedingly arranged n times, we write R^n (or L^n). In this way, we associate a sequence $\cdots L^{n_0}R^{n_1}L^{n_2}R^{n_3}\cdots$ to the directed geodesic line $L(x,y)$. Series [7] called it the cutting sequence of $L(x,y)$. If $L(x,y)$ starts from a vertex of some triangle, then the cutting sequence is finite on the left side. If $L(x,y)$ ends in a vertex of some triangle, then the cutting sequence is finite on the right side. Since each element of G is orientation preserving, labels R and L are invariant under G. For simplicity, we define numbers of even order always denote

L and numbers of odd order always denote R and we write the cutting sequence $<\cdots n_{-1}, n_0, n_1, n_2,\cdots>$. Series [7] showed the following theorem.

Theorem A. Let x be in $[-1\ 0)$ and y be in $[1\ \infty)$. Then the cutting sequence of $L(x,y)$ is of the form

$$<\cdots n_{-1}, | n_0, n_1,\cdots>,$$

where the symbol $|$ corresponds to the position where $L(x,y)$ and the imaginary axis cross, if and only if

$$x = -[0, n_{-1}, n_{-2},\cdots] \text{ and } y = [n_0, n_1, n_2,\cdots].$$

3. Proof of Theorem 1.

The modular group G is generated by $\tau(z) = -1/z$ and $\sigma(z) = z+1$. If $x = -[n_0, n_1, n_2,\cdots]$, $n_0 \neq 0$ then $\sigma\tau(x) = [1, n_0, n_1,\cdots]$. If $x = -[0, n_1, n_2,\cdots]$ then $\tau(x) = [n_1, n_2,\cdots]$. If $x = [0, n_1, n_2,\cdots]$ then $\sigma(x) = [1, n_1, n_2,\cdots]$. Hence we consider only the case $x>1$. It is well-known (e.g. [1]) that the arbitrary directed geodesic line except for $\{L(g(0),g(\infty))|g\in G\}$ is equivalent under G to some directed geodesic line $L(\theta_1,\theta_2)$ where θ_1 is in $[-1\ 0)$ and θ_2 is in $[1\ \infty)$. Since the point -1 is an intransitive point, Theorem 2 implies that $L(-1,x)$ is a transitive geodesic line if and only if x is a transitive point. From the above fact, we prove the following theorem for the proof of Theorem 1.

Theorem 1'. Let the continued fraction of x be of the form $[n_0, n_1, n_2,\cdots]$, $n_0 \neq 0$. Then the directed geodesic line

$L(-1,x)$ converges to an arbitrary directed geodesic line $L(\theta_1,\theta_2)$, $\theta_1 \in [-1\ 0)$ and $\theta_2 \in [1\ \infty)$ if and only if, for an arbitrary finite sequence $a_0, a_1, \cdots a_m$, where a_i is a positive integer, there exists k such that $n_k = a_0, n_{k+1} = a_1, \cdots n_{k+m} = a_m$.

Remark. The later condition implies the condition that, for an arbitrary finite sequence $a_0, a_1, \cdots a_m$ and for an arbitrary integer i, $0 \leqq i \leqq m$, there exists u such that $n_{2u-i} = a_0$, $\cdots n_{2u} = a_i$, $n_{2u+1} = a_{i+1}, \cdots n_{2u+m-i} = a_m$ (see [1]).

Proof. Since irrational numbers are dense in $\hat{R}$, it is sufficient to consider the case that θ_1 and θ_2 are irrational. Set $\theta_1 = -[0, a_{-1}, a_{-2}, \cdots]$ and $\theta_2 = [a_0, a_1, a_2, \cdots]$. From the theory of Diophantine approximations, we see, for an arbitrary $\varepsilon > 0$, there exist positive integers t and s such that

$$|\theta_1 + [0, a_{-1}, a_{-2}, \cdots a_{-t}, \omega_1]| < \varepsilon/2 \quad \text{and}$$
$$|\theta_2 - [a_0, a_1, a_2, \cdots a_s, \omega_2]| < \varepsilon/2,$$

where ω_1 and ω_2 are arbitrary numbers greater than 1. We consider the finite sequence $a_{-t}, \cdots a_{-1}, a_0, a_1, \cdots a_s$. By the assumption and the fact we remark, there exists u such that $n_{2u-t} = a_{-t}, \cdots n_{2u-1} = a_{-1}, n_{2u} = a_0, n_{2u+1} = a_1, \cdots n_{2u+s} = a_s$. Hence the cutting sequence of $L(-1,x)$ is of the form

$$\langle 1, | n_0, n_1, \cdots n_{2u-t-1}, a_{-t}, \cdots a_{-1}, a_0, a_1, \cdots a_s, n_{2u+s+1}, \cdots \rangle.$$

There exists an element $g \in G$ which maps the side of the tessera which the segment corresponding to a_{-1}, a_0 crosses to the imaginary axis. Hence the cutting sequence of $g(L(-1,x))$ $= L(g(-1),g(x))$ is of the form

$$<1, n_0, n_1, \cdots n_{2u-t-1}, a_{-t}, \cdots a_{-1}, | a_0, a_1, \cdots a_s, n_{2u+s+1}, \cdots>.$$

By Theorem A, we have

$$g(-1) = -[0, a_{-1}, \cdots a_{-t}, n_{2u-t-1}, \cdots n_1, 1] \quad \text{and}$$

$$g(x) = [a_0, a_1, \cdots a_s, n_{2u+s+1}, \cdots].$$

Therefore we have

$$|g(-1)-\theta_1| + |g(x)-\theta_2| < \varepsilon.$$

This shows that $L(-1,x)$ converges to $L(\theta_1,\theta_2)$.

To show the converse direction, we follow the above argument conversely. q.e.d.

4. An application.

Let $x = [n_0, n_1, n_2, \cdots]$ be an irrational number. We call x of constant type if there exists a constant M such that $n_i < M$ for all i (see [3]). By Theorem 1, numbers of constant type are intransitive under G. We assume that $n_0 > 0$. We consider the directed geodesic line $L(-x,x)$. The cutting sequence of $L(-x,x)$ is of the form

$$<\cdots n_2, n_1, 2n_0, n_1, n_2, \cdots>.$$

We set the element of G

$$g(z) = \frac{rz+s}{qz+p}.$$

The geodesic line $g(L(-x,x)) = L(g(-x),g(x))$ is a semicircle whose center is in $\hat{R}$, and whose diameter is

$$|g(x)-g(-x)| = \frac{2x}{q^2|x-p/q||x+p/q|}.$$

On the other hand, the cutting sequence $<\cdots n_2, n_1, 2n_0, n_1, n_2, \cdots>$ implies that $g(L(-x,x))$ cuts at most $2M$ axes which are parallel to imaginary axis and whose endpoints are integers. Hence we have

$$|g(x)-g(-x)| < 2M+2.$$

Therefore, this inequality is satisfied if and only if the inequality

$$|x-p/q| > c/q^2 ,$$

where c is a constant which is independent of p and q, is satisfied.

Next, we consider the Riemann surface H/G. The fundamental region of the modular group is $F = \{z = x+iy \,|\, 0 \leqq x < 1,\ x^2+y^2 \geqq 1\ (0 \leqq x \leqq 1/2),$ and $(x-1)^2+y^2 > 1\ (1/2 < x < 1)\}$. We identify the Riemann surface H/G with this fundamental region. By π we denote the natural projection from H to H/G. All the elements of the set $\{L(g(-x),g(x)) \,|\, g\in G\}$ exist below the line $y = M+1$ if and only if the geodesic line $\pi(L(-x,x))$ on H/G is in $F\cap\{z = x+iy \,|\, y \leqq M+1\}$. Hence we conclude the following theorem.

Theorem 3. The following three conditions are equivalent.

i) x is of constant type.

i i) $|x-p/q| > c/q^2$ for all integers p and q which are relatively prime numbers, where c is a constant which is independent of p and q.

iii) $\pi(L(-x,x))$ is in some compact set in H/G.

Remark. The equivalence of i) and ii) has been already proved by other method (e.g. [3]).

References.

[1] E. Artin, Ein mechanisches System mit quasiergodishen Bahnen, Abh. Math. Univ. Hamb., 3 (1924), 170-175.

[2] E. Hopf, Ergodentheorie, Ergebniss der Mathematik, Band 5, no.2, Springer-Verlag, Berlin, 1937.

[3] S. Lang, Introduction to Diophantine Approximations, Addison-Wesley Publishing Company, 1966.

[4] J. Lehner, Discontinuous Groups and Automorphic Functions, Mathematical Survey, no.8.

[5] P. J. Myrberg, Ein Approximationssatz für die Fuchsshen Gruppen, Acta Math., 57 (1931), 389-409.

[6] P. J. Nicholls, Transitivity properties of Fuchsian groups, Can. J. Math., 28 (1976), 805-814.

[7] C. Series, The modular surface and continued fractions, J. London Math. Soc.(2), 31 (1985), 69-80.